1999

Forecast Performance of Threshold Autoregressive Models - A Monte Carlo Study

Abdurrahman Bekir Aydemir

ISSN:0318-725X
ISBN:0-7714-2213-X

RESEARCH REPORT 9910

Forecast Performance of Threshold Autoregressive Models - A Monte Carlo Study

by

Abdurrahman Bekir Aydemir

November 1998

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario, Canada
N6A 5C2
econref@julian.uwo.ca

Forecast Performance of Threshold Autoregressive Models - A Monte Carlo Study*

Abdurrahman Bekir Aydemir

The University of Western Ontario

Social Sciences Center

Department of Economics

London, ON, Canada

N6A 5C2

aaydemir@julian.uwo.ca

+ (519) 661 2111 ext. 5239

November, 1998

*I am grateful to John Knight for his guidance, advice and continuous support. I would also like to thank seminar participants at the University of Western Ontario, especially R.A.L. Carter for helpful discussions. Usual disclaimers apply.

Forecast Performance of Threshold Autoregressive Models - A Monte Carlo Study

Abstract

Threshold Autoregressive Models (TAR) along with other nonlinear time series models have attracted much attention in recent years in time series analysis. TAR models have been applied to a variety of time series. It has been reported that they have a good in sample fit but like many other non-linear time series models cannot improve out of sample forecast performance. Within a controlled simulation framework, we study the forecast performance under two types of non-linearity: shift in the mean and shift in the volatility of the process. We illustrate that estimation of the lag parameter and threshold value are crucial for forecast performance. Monte Carlo results show that TAR model performs much better than a Random Walk (RW) model; however, it provides no significant improvement over a linear Autoregressive (AR) model. Conclusions on the relative forecast performance of TAR models based on a single data set can be quite different than long run (Monte Carlo) results.

Key Words: Nonlinear Time Series, Threshold Model, Forecast Performance, Monte Carlo

1. Introduction

Linear time series models have been used extensively in time series analysis and they have had a considerable level of success over the years. However, as pointed out by many practitioners of time series analysis, linear models cannot mimic some of the observed features in many time series data. The issues of asymmetric volatility, time irreversibility, limit cycles, sudden bursts at irregular time intervals etc. lead to a new class of models referred to as non-linear time series models.

The threshold autoregressive (TAR) model, first introduced by Tong (1978), is one of the nonlinear models discussed in the literature. The major features of this class of models are limit cycles, amplitude dependent frequencies, and jump phenomena. Tong and Lim (1980) show that the threshold model is capable of producing asymmetric, periodic behavior exhibited in the annual Wolf's sunspot and Canadian lynx data.

TAR models have been extensively applied to diverse fields, ranging from water pollution (Tong, 1990, p. 278) to stock market returns and exchange rates. There are many studies available in the literature which compare in sample and out of sample forecast performance of these models with linear and other nonlinear models (e.g. ARCH and bilinear models) using different data sets. Like most non-linear techniques TAR models give a good in sample fit. The results about the out of sample forecasting, however, are mixed. While some studies conclude that TAR models have a poor performance in out of sample forecasting, others show that TAR models are superior when compared to random walk (RW) model, autoregressive (AR) model and other nonlinear models.

In their review of the recent developments in non-linear time series modelling, testing, and forecasting, De Gooijer and Kumar (1992) conclude that no uniformity seems to exists in the evidence presented on the forecasting ability of non-linear models. They point out the need for research, in a controlled simulation framework, to clarify this issue.

The aim of this paper is to study the factors affecting the forecast performance of TAR models in a Monte Carlo framework. We study two kinds of nonlinearities: shift in the mean of the process and shift in the volatility. We raise the question "Are there distinguishable patterns in the data affecting the forecast performance of TAR models which may lead to different conclusions in the literature?". Moreover, we link the simulation results to the estimation problems inherent in the TAR modeling. Data is generated from self-exciting TAR (SETAR) models. We use both empirical and estimated SETAR models as Data Generating Processes (DGP's). For the latter, we use exchange rate series. Nonlinearity is reported in exchange rate data by various studies in the literature.

The conclusions of our study are: the estimation of the lag parameter and threshold value are crucial for forecast performance. Increasing the nonlinearity of the process, by increasing the magnitude of the shift in the mean or volatility, causes the estimates of threshold value to deteriorate. Therefore, increasing nonlinearity through missepecification of the threshold lag or value, causes the forecast performance to deteriorate significantly. Also, long run forecast performance results obtained from repeated use of a TAR model can be quite different from forecast comparisons based on a single data set. Depending on the measure used for forecast comparisons the results can also differ.

In the next section we give the motivation for TAR models. Section III reviews the literature that assess the forecast performance of TAR models. The estimation method used in the Monte Carlo study is presented in section IV, while section V presents the Monte Carlo experiments and discusses their results. Section VI concludes.

2. TAR Models

A TAR model allows the parameters of a linear (AR) model to vary according to the values of a finite number of lags in some conditioning variable, z_t (threshold variable). TAR models may also be regarded as "piecewise linear" approximations to more general kth-order non-linear autoregressions of the form $y_t = f(y_{t-1}, y_{t-2}, ..., y_{t-k}) + e_t$. However, it is piecewise linear in the space of the threshold variable, not piecewise linear in time. The non-linearity can arise in conditional mean, conditional variance or both.

The "Self-exciting" TAR (SETAR) model is a subset of TAR models where the conditioning variable is assumed to be the dependent variable itself. A simple SETAR model of a time series y_t with threshold variable y_{t-d} is defined as follows, where d is a positive integer and is referred to as the threshold lag:

$$Y_t = \phi_0^{(i)} + \phi_1^{(i)} Y_{t-1} + ... + \phi_p^{(i)} Y_{t-p} + \varepsilon_t^{(i)} \text{ if } Y_{t-d} \in R_i, \quad i = 1, 2, ..., k \qquad (1)$$

where the R_i's form the non-overlapping partition of the real line i.e. $\cup_{i=1}^{k} R_i = R$ and $R_i \cap R_j = \emptyset$ if $i \neq j$, k is the number of threshold regimes, p is the AR order (which can vary across regimes) and $\left\{\varepsilon_t^{(i)}\right\}_{t=0}^{T}$ is a sequence of i.i.d. normal random variables with mean zero and variance σ_i^2 such that $\left\{\varepsilon_t^{(i)}\right\}$ and $\left\{\varepsilon_t^{(j)}\right\}$ are independent if $i \neq j$. In a SETAR model with k regimes, there exists $(k-1)$ threshold values $(r_1, ..., r_{k-1})$ which determine the k non-overlapping partition of the real line $(R_1, ..., R_k)$. Model (1) reduces to a linear AR process when $k = 1$. Since we will focus on SETAR models, in the rest of the text we will refer to a SETAR model simply as a TAR model.

Tong (1990) discusses that, the basic idea of a 'threshold' is the local approximation over the states, i.e. the introduction of regimes via threshold. This idea is called the 'threshold principle' where a complex stochastic system is decomposed into simpler subsystems. Priestly (1965), Priestly and Tong (1973) and Ozaki and Tong (1975) use a similar idea in the analysis of nonstationary time series and time dependent systems. Local stationarity in those papers is an analogue of local linearity in threshold models. Inherent in the threshold models is a feedback mechanism from the past observations to the parameters. Such a mechanism is lacking for the Markov chain-driven models, where the switches are supposed to be driven by external events.(For a review of threshold models and various issues related to their estimation, testing, forecasting etc. see Aydemir (1996)).

3. Nonlinearity and Forecast Improvements

There is wide agreement that a variety of high-frequency asset returns are well described as linearly unpredictable, conditionally heteroscedastic, and unconditionally leptokurtic (Diebold and Nason, 1990). Cootner (1964) and Fama (1965) are the early studies documenting linear unpredictability. There is also an apparent volatility clustering in asset returns (Engel (1982)). It is agreed in the literature that many time series of asset returns, while approximately uncorrelated, are not temporally independent. The dependence arises through persistence in the conditional variance and perhaps in other conditional moments (Diebold and Nason, 1990). Many empirical studies e.g. Domowitz and Hakkio (1985), Engle, Lillien and Robins (1987), among others, detect significant nonlinearity in conditional means of various asset prices and other aggregates. Scheinkman and LeBaron (1989) find strong evidence of nonlinearity in common stock returns. They suggest that this property can be exploited for improved point prediction.

Studies on foreign exchange rates find that the series are $I(1)$ (integrated of order one) processes and that changes of exchange rates are uncorrelated over time. Therefore, changes in the exchange rates are not linearly predictable in general (Kuan and Liu, 1995). Westerfield (1977), Boothe and Glassman (1987), and Diebold and Nerlove (1989) provide extensive reviews of linear unpredictability and leptokurtosis in exchange rates. Hsieh (1989) finds that changes of exchange rates may be nonlinearly dependent, even though they are linearly uncorrelated. There is also a lot of evidence on conditional heteroscedasticity in the form of ARCH. Gallant, Hsieh and Tauchen (1988) and Hsieh (1989) report evidence of residual nonlinearity in exchange rates, after controlling for conditional heteroscedasticity. These results suggest that possible existence of nonlinearity in conditional means may be exploited to improve predictability.

TAR models has been mainly used for exchange rate predictions, modelling volatility series of stock returns, interest rates and macroeconomic time series such as GNP.

Although there is ample evidence in the literature on nonlinearity in exchange rates, with the exception of a few studies, most authors report that predictions of nonlinear models are not superior to the predictions of a RW. Diebold and Nason (1990, p.317) offer a number of possible explanations for this puzzle.

Peel and Speight (1994) try to exploit the nonlinearity in weekly spot exchange rates using an ARCH model and compare the in sample performance with TAR, bilinear and linear models. They report that a TAR model gives the best prediction in one of the series, but other models are superior to TAR in the other two series. Chapell et al. (1996) use a TAR model for modeling daily French franc/Deutschemark exchange rate. Their result suggests that a TAR model provides a better explanation and superior forecasting performance than a RW. They also note that TAR models could not improve over a RW in other exchange rate series they explored. The authors fit their models to the log transformed series which they find to be integrated of order (1).

Dacco and Satchell (1995) assume a RW model in each regime in a two regime threshold model, and derive the necessary and sufficient conditions for MSE(TAR) to be less than that of the RW in terms of the probability of misspecifying the regime. They assume that the parameters of the RW models in each regime are known. They fit a model to DM/Dollar exchange rate and simulate the estimated model to get estimates of the threshold value in each replication of the process. They find that the prediction of the threshold value doesn't satisfy their necessary and sufficient condition 70 % of the time. They conclude that "It requires only a small misclassification when forecasting which state the world will be in, to lose any advantage from knowing correct model specification".

Chan and Cheung (1994) study forecasting performance of an extension of generalized-M (GM) estimates to nonlinear case and compare with the TAR process. Their data is the monthly average wholesale prices of regular leaded gasoline in the U.S.. They find that both estimation techniques provide improved mean absolute error (MAE) and RMSE over a linear AR model. Tiao and Tsay (1994) use the series of quarterly US real gross national product and show that the TAR model not only provides better post sample multi step ahead forecast based on MSE criterion, but also highlights various interesting features of the GNP series. Byers and Peel (1995) use monthly industrial production series for six countries. Comparing AR, bilinear and TAR model, no model consistently dominates when MSE is employed as the forecasting criterion. Cao and Tsay (1992) use monthly volatility series of stock returns and explore the out of sample forecasts for TAR, linear ARMA and nonlinear GARCH and EGARCH models. They find that TAR models consistently outperform the linear ARMA models in multi step ahead forecasts for large samples. TAR models provide better forecasts than the GARCH and EGARCH models for the volatilities of large stock returns. EGARCH models give the best long-horizon volatility forecasts for small stock returns. The present author in a previous paper (Aydemir, 1996) compares the forecast performance of TAR, AR and RW models using four short and long US interest rates. It is found that TAR models have a better in sample and out of sample forecast performance compared to other models. Moreover, the regime shifts indicated by the threshold models are matched with Federal policy changes reported in the literature.

Clements and Smith (1996), consider SETAR models proposed in the literature for various economic time series (exchange rate data, U.S. GNP series, U.K. savings ratio, Canadian lynx data, Wolf's sunspot numbers). Based on these estimated empirical models in the literature, they design a Monte Carlo study by generating data from the estimated models. They compare the forecast performance of SETAR models with the linear ($AR(p)$) models and the RW model based on mean-square forecast errors. In their Monte Carlo study they take two approaches. First, they assume the SETAR model is the SETAR DGP (i.e. they assume that the number of regimes, the threshold values, the delay lag, the orders of the process in each regime and the model coefficients are all known). In the second approach, they assume that the model is the DGP up to AR parameters being unknown. In this latter approach, they run another set of experiments where AR orders in the regimes are not known and are selected by AIC (Akaike Information Criterion), conditional on the number of regimes, threshold value, delay lag and distribution of disturbances being known. Monte Carlo results from the first approach suggest little evidence that SETAR models of exchange rate provide better multi-period forecasts than linear models. For U.S. GNP, however, SETAR performs better. They also have evidence that SETAR model is superior to a RW model in forecasting. Compared to the RW model they find positive gains (in terms of a reduction in MSE). For U.S. GNP, for example, MSE of RW divided by the MSE of the SETAR model for 1-step-ahead forecasts is 1.43 and this declines to 1.04 for 3-step ahead. The same ratio for 1-step-ahead is 10.86 for Lynx data, 1.53 for sunspot numbers. Using the second approach (by estimating AR coefficients) results in reductions in these gains. Finally, in empirical examples, they find a rejection of the null hypothesis that the regimes are serially independently distributed is necessary but not sufficient for an improved forecast performance.

In this paper we focus on forecast uncertainty emanating from parameter estimation and model selection. In our Monte Carlo study all the parameters of the threshold model- except the number of regimes which is set equal to two- are estimated from the data. The data is generated in two ways: In the first approach we assume an empirical DGP in the form of a SETAR model with two regimes. In the second approach SETAR models are fit to exchange rate data and the estimated empirical models are used as DGP's. This approach allows for a controlled simulation framework where the empirical relevance of the results are supported with the results from the data generated by the estimated empirical models. The alternative models are not only compared in terms of MSE but also in terms of other measures e.g. Mean Absolute error, sign prediction etc. The gains we find are much more significant than those reported by Clements et al.

Having presented the various studies using TAR models trying to achieve forecast improvements, in the next section we will present the estimation method we used in estimation of TAR models in this study.

4. Estimation Method

For TAR models we use the modelling procedure of Tsay (1989). Using the concepts of 'arranged autoregression' and 'local estimation' a TAR model is transformed into a regular change-point problem. Local estimation means that a sequential estimation is performed with a fixed-length window that controls the number of observations used. An arranged autoregression uses the magnitude of the threshold variable Y_{t-d} , rather than using the time index, to control the data flow of the window. Using arranged autoregression the threshold problem is transformed into a switching regression problem for which statistics can be derived to test for model changes and to explore the dynamic structure of the process.

Cao and Tsay (1992) illustrate the idea of arranged autoregression using a simple example. For purposes of illustration their example is reproduced below. Suppose that we have a TAR model with two regimes (TAR(2)) as follows:

$$Y_t = \begin{cases} 1.3Y_{t-1} - 0.4Y_{t-2} + \varepsilon_t^{(1)} & if\ Y_{t-2} < 0 \\ 0.3Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t^{(2)} & otherwise \end{cases} \tag{2}$$

where $\varepsilon_t^{(i)}$ are independent sequences of i.i.d. $N(0,1)$ random variables, $L_1 = (-\infty, 0)$ and $L_2 = [0, \infty)$ are the two regimes obtained from partition of the real line by the threshold value of zero. Part (a) of Table 1 (see appendices for the tables) shows the usual set up for fitting an AR(2) model to 10 consecutive observations of the TAR(2) model. Such a set up cannot yield consistent parameter estimates. In Part (b) of the table arranged autoregression rearranges the observations according to Y_{t-2}. The first four rows belong to the first regime whereas the last four belong to the second regime. Consistent estimates of the variables can be obtained once the observations are rearranged.

TAR modelling procedure consists of the following steps (See Tsay (1989) for details):

1. Select a tentative AR order p and a set of possible threshold lags S.

2. For a given p and every element d of S fit an arranged autoregression, perform the threshold nonlinearity test $\widehat{F}(p,d)$.(1)

3. If nonlinearity is detected, select the threshold variable Y_{t-d} based on the test results of step 2. (The delay parameter which yields highest F-statistic is chosen as the delay parameter of the model.)

4. Perform an arranged autoregression to locate the possible threshold values. (2)

5. Estimate the specified TAR model by conditional least squares.

6. Check the estimated TAR model and refine it if necessary.

In the above algorithm the F-statistic is used for both detecting nonlinearity and for choosing the delay parameter d. Several tests have been proposed in the literature to detect the potential nonlinearities in the data. Simulation studies reveal that none of the tests performs best overall. Therefore it is not uncommon in the literature to use four or five different tests in model selection. However, in a Monte Carlo framework it is not possible to calculate different tests as this will increase the computation time enormously. After detecting nonlinearity and determining d, the algorithm locates the possible thresholds. This is usually done by visual inspection of the scatter plot explained in the footnote of Step 4. To automate the above algorithm for finding the threshold value we follow the following strategy. After fitting an arranged autoregression for a given p and d, we construct a confidence interval around the initial standardized residuals assuming linearity. Then, any outlier which falls outside the confidence band is treated as a potential threshold value. We tested the above strategy and found that the threshold values found and the AR models fitted at either side of the regime are reasonably close to the values that can be obtained by making a visual inspection of the data. For a single data set visual inspection is preferred in detecting the potential threshold values. However it is also obvious that this method cannot be used in a simulation framework with thousands of replications.

The final step of the procedure involves refinement of the fitted model. This step is left out of our algorithm as it requires subjective decisions and is computationally too costly.

The AR order p is selected by considering the partial autocorrelation function (PACF) of Y_t. Tsay (1989) notes that PACF can be preferred over some information criterion such as the Akaike information criterion (AIC) because (a) PACF often provides guidance for a reasonable value of p, (b) the information criterion could be misleading when the process is indeed nonlinear, and (c) the AR order can be refined, if desired, at Step 4.

5. Monte Carlo experiments

In the Monte Carlo experiments we study two kinds of nonlinearities in the TAR framework. We concentrate on models with two regimes for simplicity. The nonlinearities examined are: nonlinearity in the mean (i.e. each regime has a different mean) and nonlinearity in the volatility (i.e. each regime has a different volatility). There is no reason why both of these nonlinearities would not arise simultaneously in a real data set. For clarity of the arguments, however, each nonlinearity is treated separately.

For each data set we compare in sample and out of sample forecast performance of TAR, linear AR and RW models. As discussed before RW seems to challenge the linear and nonlinear models in forecasting time series where nonlinearity is detected, especially in exchange rate predictions. We do not estimate ARCH or other nonlinear models such as a bilinear model since the focus of this study is the performance of TAR models rather than a comparison of all available linear and nonlinear alternatives. Cao and Tsay (1992) explore the use of TAR models in describing monthly volatility series. They present a rich set of results where out of sample forecasts are used to compare the TAR models with linear ARMA models and nonlinear GARCH and EGARCH models. Interested readers are referred to this study.

For evaluation of alternative models in our study we report different measures including MSE, MAE, mean error (ME) and sign predictions. The most frequently used measure by econometricians is the MSE which is the principal criterion for comparing forecasts. Edison (1991) notes that MAE is a useful criterion when the exchange rate distribution has fat tails, even if the variance is finite. The ME provides another measure of robustness. Also by comparing the MAE and ME it is possible to ascertain whether a model systematically over- or underpredicts.

In "sign predictions" the criterion is whether a model predicts the direction of the change of series correctly. Leitch and Tanner (1991) find that the direction of change criterion is the best proxy among several (including MSE and MAE) for choosing forecasts of interest rates on their ability to maximize expected trading profits. Engel (1994) notes that the direction of change criterion may be the right criterion for maximizing the welfare of the forecaster under certain conditions. Under pegged exchange rate systems, central banks are often interested in the direction of changes in the exchange rate. They might need to intervene to support the currency if it is expected to depreciate, regardless of the size of the expected depreciation. Levich (1981) and Merton (1981) also report that financial economists are usually interested in sign predictions which yield important information for financial decisions such as market timing.

Simulation results report all of the above statistics. We concentrate on the MSE measure during our discussions as this is the measure most often reported in the literature.

5.1. Shifts in the mean of the process

In this experiment, we keep the volatilities of the regimes constant and study the effect of increasing or decreasing the distance between the means of the processes on the estimation and forecast performance. Let u denote the regime with the higher mean (upper regime) and l the lower regime. The model we use in the Monte Carlo experiment is:

$$Y_t = \begin{cases} \phi_0^u + \phi_1^u Y_{t-1} + \varepsilon_t^u & if\ Y_{t-1} > r \\ \phi_0^l + \phi_1^l Y_{t-1} + \varepsilon_t^l & if\ Y_{t-1} \leq r \end{cases} \quad (3)$$

where ε_t^i have zero mean and constant variance across regimes. If we denote the mean of regime i by $\mu(Y_t^i)$ and the volatility by $\sigma(Y_t^i)$, the experiment aims to analyze the effect of $(\mu(Y_t^u) - \mu(Y_t^l))$ on the forecast performance of the threshold model relative to the linear AR and RW models, where we impose the restriction $\sigma(Y_t^u) = \sigma(Y_t^l)$. We make the assumption of equal volatility since volatility can potentially affect the estimation and forecast performance which we study in the next experiment. The mean of either regime is chosen to be of equal distance to the threshold which assures that on the average the process stays in either regime for equal amounts of time. By the design of the experiment, it is clear that as $(\mu(Y_t^u) - \mu(Y_t^l))$ increases, the frequency of shifts between the regimes will decline while the average amount of time spent in either regime will increase. As $(\mu(Y_t^u) -$

$\mu\left(Y_t^l\right)) \longrightarrow 0$, model (3) reduces to a simple linear $AR(1)$ process. Therefore, we refer to an increase in $(\mu(Y_t^u) - \mu\left(Y_t^l\right))$ as an increase in "overall nonlinearity" of the process.

For each experiment we generate 1000 replications of the process each with 750 observations. The first 250 observations are discarded to eliminate start-up effects. For the remaining sample size of 500 observations, the first 495 observations are used for estimation while the remaining 5 are kept for forecasting. We imagine a financial analyst who gets data on daily exchange rates from Monday to Friday. The analyst reestimates a new model with the new observations, on the last day of every week, keeping the window size of estimation constant, which is used for forecasting the following week. Therefore each model estimated gives us 5 $1 - step - ahead$ forecasts, 4 $2 - step - ahead$ forecasts etc. and finally 1 $5 - step - ahead$ forecast. The parameters to be estimated are the delay parameter d, threshold value r, order of linear AR models in either regimes p and q, and finally the AR coefficients in both regimes.

To compare the results from each experiment we use two summary statistics. The first measure is the simple mean, e.g. in the case of a threshold model and MSE as the measure, the mean of MSE's over all replications:

$$\overline{MSE}(TAR) = \sum_{i=1}^{1000} MSE^i(TAR)/1000 \qquad (4)$$

If the assumption is that the individual has a quadratic loss function, then, MSE becomes the relevant criterion, and the measure in (4) gives the average loss an individual will realize.

The second summary statistic we report is:

$$S_2 = \left(\frac{\sum_{i=1}^{1000}\left[\left(MSE\left(TAR\right) - MSE\left(RW\right)\right)/MSE\left(RW\right)\right]_i}{1000}\right) * 100 \qquad (5)$$

This statistic gives a measure of the average percentage improvement (or decay) in forecast performance of the threshold model with respect to RW. We report S_2 only for MSE.

The percentage of correct sign predictions are reported by $\overline{SP}$.

Tables (2) to (4) give the parameters of the data generating processes along with some descriptive statistics. **(3)**

As it is clear from the tables, the distance between the two regimes increases moving from Table 2 to Table 4. Notice that ϕ_1 remains constant along the three simulations implying that past observations' effect on the current level of the series Y_t is the same in all experiments. The last two columns give the mean and the volatility of the linear AR models in each regime.

Table 5 reports the results of experiment 1 for 1-step-ahead, out-of-sample forecasting. We run two tests for significance of forecast differences between different models and along different simulations. See tables 29-32 in appendix for test results. The tests are discussed before the table 29 in appendix. Footnote 4 discusses the large standard errors reported for S_2 .

Tables 6 and 7 report the results for 3-step-ahead and 5-step-ahead out of sample forecasting (Figures within paranthesis are standard errors). We don't report in-sample forecast performance results since the results are qualitatively the same as out-of-sample.

As we move from simulation 1 to simulation 3, the performance of threshold and AR models gets worse in terms of MSE, MAE and sign predictions. For RW model, however, there is an improvement in the forecast performance in terms of the first two criterion.

The increase in $\overline{MSE}(AR)$ is not surprising. The underlying assumption of an AR model is the 'linearity' of the process. As the nonlinearity becomes stronger when the means of the regimes get more distant from each other, the misspecification of the model becomes more severe, leading to worse estimates.

It is interesting to observe that $\overline{MSE}(RW)$ decreases from simulation 1 to 3. RW model always estimates that the world will be in the current regime in the next period. As long as there are no shifts between the regimes, the MSE of a RW model arises from the volatility of the linear AR model in each regime. Existence of shifts between regimes adds to MSE of RW due to wrong predictions of the regimes. The more apart the processes are at the either side of the threshold, the more severe is the error from the misprediction of the regime. As we move from simulation 1 to 3, however, the frequency of regime shifts decreases. In the above experiment from these two opposing effects the second effect dominates, leading to decline in MSE.

In each simulation above the process stays, on average, an equal amount of time in either regime. With the decreasing frequency of regime shifts, once the process is in one regime it stays there longer. This has an effect on the estimation of parameters of each locally linear AR model in an arranged autoregression framework, illustrated as follows:

a) Suppose the process remains only one period in a given regime and shifts to the other one in the next period. Also, assume that we are estimating an $AR(1)$ model in each regime. In an arranged autoregression framework discussed previously, this results in a vector of dependent variables $\mathbf{X}_t$ coming from regime 1, while the vector of explanatory variables $\mathbf{X}_{t-1}$ coming from the second regime.

b) Suppose, instead, that the process stays on average 100 periods in a given regime once it is there. If we call $[X_t, X_{t-1}]$ as a 'case' to be used in estimation of the locally linear $AR(1)$ model, only 1 case out of 100 will include an explanatory variable from the other regime.

From Table 5, we noted that the performance of the AR model gets worse as the nonlinearity of the process increases. In the estimation of a threshold model, we make the assumption of 'local' linearity in each regime. Due to points (a) and (b) above, this 'local linearity' assumption is a strong one when shifts between regimes are frequent. It is a weaker assumption, however, if shifts are less frequent and the process stays a long time once it is in a regime. This will effect the estimation of parameters of the locally linear models in a threshold model.

A second, and perhaps more important factor that affects the estimation of a threshold model is the estimation of the threshold value which we will discuss in more detail in the next section. Misspecification of the threshold variable or the threshold value will also lead to inconsistent estimates of the other parameters of the model. As we will show later, when we move from simulation 1 to 3, i.e. when the frequency of shifts declines, the estimation of threshold variable and the threshold value become poorer. As a result of the above mentioned effects we observe the increasing MSE of the threshold model. Notice also that in a RW model, as long as there are no shifts in the regime, RW will estimate the next periods' regime correctly. However, even if there is no shift in the regime, a threshold model may misspecify the next period's regime due to a wrong prediction of the threshold variable or the level of the process.

The most important observation from Table 5 is the performance of the threshold model relative to RW model. The summary statistic S_2 shows that as we move from simulation 1 to 3, the average percentage reduction in MSE by threshold relative to RW declines. Also the $\overline{MSE}(TAR)/\overline{MSE}(RW)$ ratio increases from simulation 1 to 3, by (0.6418, 0.7048 and 0.7655) respectively. Similar results hold for $\overline{MAE}$ as well.

The performance of the threshold model is quite close to that of the AR model as can be seen from the table in the first two simulations. However, the TAR model becomes superior in simulation 3.

Finally it is important to point out the performance in terms of sign predictions. Since RW predicts that the value of the process next period will equal today's value, it trivially has a zero correct sign prediction rate. Although declining along the simulations, this rate is positive for both threshold and AR models.

Two important results emerge from the above discussions. First, the values reported in the tables are averages over 1000 replications, and therefore, give an idea of the average gain (loss) of using a model relative to an alternative model. While, in certain replications of the process RW outperforms the other two alternatives, this doesn't hold on average. Most of the results reported in the literature look at a certain data set and compare the relative performance of the models using that particular data set. One cannot, however, draw general conclusions from one particular data set, thus emphasizing the importance of making a rich analysis by using different data sets. Also, within a single data set the use of different periods with equal window sizes should be examined before reaching a conclusion about the relative performance of any model.

Secondly, we see that it is important to understand what the loss function of a financial analyst is before we can conclude whether a method is superior to another. All the above summary statistics on forecast performances are obtained with different loss function assumptions. Under the hypothetical model of simulation 3, if the loss function implies a quadratic relation, then MSE becomes the relevant criterion, and the threshold model will be preferred to a RW. We reach the same conclusion if sign predictions are the relevant criterion. However, if ME is implied by the loss function, then, RW is superior.

In the following section we look at the estimates of the threshold variables and their effect on forecast performance more closely. It is agreed in the literature that threshold variable estimation is the hardest part of estimation in a threshold model given the assumption of the number of regimes. Since most studies use real data sets where the underlying data generating process is unknown, it is not possible to see the effect of wrong threshold estimations on the forecast performance and overall fit of the model.

5.1.1. Threshold estimation

In this section, we first report the estimation results in the above three simulations in terms of threshold variable predictions. Then, we assume that the true threshold variable (threshold lag) and the value of the threshold is known. As a result of this assumption the estimation of the threshold model becomes trivial. All that is needed is to sort the observations according to the threshold variable and estimate a locally linear AR model for each regime. This gives us an idea of the effect of misspecification of the threshold parameter on the forecast performance. We will again discuss our results in terms of 1-step-ahead, out-of-sample forecasting.

Table 8, reports the descriptive statistics of the estimated threshold values in all three simulations. From Table 8, we observe that as we move from simulation 1 to 3, the misspecification of threshold estimates increases. When the two regimes are closer to the threshold variable, the range in which we search for the threshold is tighter and closer to the true value. When means get further apart from each other, however, this range increases and misspecification probability increases. We also observe in our simulation, which we don't report in the above table, that the percentage of time the threshold lag is misspecified increases as we move from simulation 1 to simulation 3. These results show that the threshold estimation becomes poorer as the nonlinearity increases.

Next, we show the effect of threshold misspecification on forecast performance of the threshold model. We repeat simulations 1 to 3 assuming that the true value of the threshold variable is known. Model specification (the order of the parameters of AR model in each regime) and forecasting are done under this assumption and these simulations are referred to as 4 to 6 respectively. Table 9 reports $\overline{MSE}$ and S_2 measures for simulations 1 to 6.

Table 9 shows that estimation of threshold becomes critical as the nonlinearity of the process increases. The gain from knowing the true value of the threshold is much less in simulations 1 and 4 as compared to other pairs. In the limit, when the distance between the means of the processes approaches to zero, the ratio of $\overline{MSE}$'s will approach to 1.

Notice that along simulations 4 to 6 there is a slight improvement in terms of $\overline{MSE}$. In these simulations, since the true value of the threshold is known, we know the correct regime the process will be in the next period. Therefore, the prediction of the next period's regime does not affect the $\overline{MSE}$ measure. We can, however, improve parameter estimates of the locally linear models as we move towards simulation 6, following from the discussions of the previous section. This leads to the improvements in the above table in terms of $\overline{MSE}$ values.

We discussed before that, in simulations 1 and 2 TAR and AR models' forecast performance were almost identical. Without presenting the results, we note that, when true value of threshold is known, TAR model performs much better than AR model in all three simulations.

The closest study in the literature to ours is that by Dacco and Satchell (1995) which is discussed in section 3.2. Although, the condition they derive for $MSE(TAR)$ to be less than the $MSE(RW)$ is a useful one, it is fairly restrictive as it assumes a RW at either side of the regime for the TAR model. They assume that the delay parameter and the parameters of the RW models in each regime are known. After estimating the threshold value, they use the true values of the RW coefficients in each regime. However, as long as the estimated threshold value is not equal to the true value of the threshold, it implies different coefficient values for the parameters of RW models in each regime than the true values. Using true values of RW coefficients is therefore not consistent with the estimation of the threshold value. Finally, although they derive the necessary and sufficient condition, they do not check if it is satisfied empirically using the simulation results. Although our results are specific to the parameters chosen for simulation, the model we use is a more general one. Threshold lag, parameters of the regimes are not restricted and are estimated in each simulation.

In all simulations above we did not have a real life motivation for the parameters of the model. In the next section we aim to illustrate that our results can arise using a real data set as well, hence the results we show are not artificial. We will illustrate only some of the results rather than repeating the whole exercise above.

5.1.2. UK/DM Exchange Rate

The data set we use is the daily UK Pound versus German Deutschemark exchange rate over the period 1 May 1990 to 30 March 1992 giving us a total of 503 observations. The data is obtained from Datastream. See appendix for a plot of the series and first differences of the natural logarithm of the series.

Many authors including Mussa (1979) and Meese and Rogoff (1983) state that natural logarithms of exchange rates should follow simple RW. If this hypothesis is true, then, it implies that the natural logarithm of the series will have a single unit root (i.e. will be $I(1)$) and first difference of the natural logarithms will be a stationary, $I(0)$ process. Let the series be denoted by X_t, $E_t = \ln X_t$ and $Y_t = E_t - E_{t-1}$. There is strong evidence that series E_t is $I(1)$ and no evidence of serial correlation in the Y_t series.

We partition the series Y_t into two subperiods with an equal window size of 483 observations, and fit a TAR model to each series. The first series (Y_t^1) is the first 483 observations while the second (Y_t^2) is the last 483 observations (i.e. for the first series we drop the last 19 observations and for the second series we drop the first 19). Table 10 reports the descriptive statistics for the original series, and the other two series we obtain Y_t^1 and Y_t^2. In Table 11 the values of statistics for test of nonlinearity are presented for up to 10 lags, which shows evidence of nonlinearity in all three series (For all lags the null hypothesis of linearity is rejected at 5 % level).

We estimate a threshold model for Y_t^1 and Y_t^2, the parameters of which are presented in Table 12, along with the mean and the volatilities of regimes implied by parameter estimates. Comparing the volatilities of the upper and lower regimes along the two series, we see that they are approximately equal. Lower regimes have a zero mean in both series, but the mean of the upper regime in Y_t^1 is about twice the mean of the upper regime in Y_t^2. In previous simulations while the means of the regimes in either side of the regime were different, we imposed the restriction of equal volatility in both regimes. It is difficult to satisfy this condition with real data sets.

We replicate each estimated model 1000 times (simulation 7 and 8 for Y_t^1 and Y_t^2 respectively). Table 13, presents the results.

In Table 13, $\overline{MSE}(TAR)/\overline{MSE}(RW)$ is equal to 0.5078 for simulation 7 and 0.4888 for simulation 8. The performance of TAR and AR models are very close to each other. Most of the changes in other statistics are as expected. For simulation 7, although the mean value of threshold estimate is closer to the true value compared to simulation 8, its standard deviation is higher.

5.2. Shifts in the volatility

This experiment considers the shifts in the volatility of the process, keeping the mean constant. Using the earlier notation the process we analyze in this section is:

$$Y_t = \begin{cases} \phi_0^u + \phi_1^u Y_{t-1} + \varepsilon_t^u & if \ |Y_{t-1}| > r \\ \phi_0^l + \phi_1^l Y_{t-1} + \varepsilon_t^l & if \ |Y_{t-1}| \leq r \end{cases} \tag{6}$$

where ε_t^i have zero mean and variance $(\sigma_\varepsilon^i)^2$ in each regime. Denoting the mean of regime i by $\mu(Y_t^i)$ and the volatility by $\sigma(Y_t^i)$, the experiment is to see the effect of $(\sigma(Y_t^u) - \sigma(Y_t^l))$ on the forecast performance of threshold model relative to the linear AR and RW models, imposing the restriction $\mu(Y_t^u) = \mu(Y_t^l)$. The difference of the process in (6) from that in (3) is that, model (6) implies a TAR model with two thresholds which are symmetric around zero, namely r and $-r$. As long as the series remains within the band defined by the threshold the process is in the lower regime. Once the band is hit, the process switches to the second regime. We assume that $\sigma(Y_t^u) > \sigma(Y_t^l)$and $\phi_i^u = \phi_i^l$. As $(\sigma(Y_t^u) - \sigma(Y_t^l)) \to 0$ the process in (6) reduces to an $AR(1)$ model.

The same sample size is used for estimation and forecasting as in previous simulation. Also, the same statistics will be reported. Tables 16 to 18 present the parameter specifications used in simulations.

As we move from Table 16 to 18, the volatility of upper regime decreases, with the limiting case of Table 18 the process reducing to an AR(1) model. Results of simulation 11 can be used to assess the performance of TAR model under a linear case. Table 19 summarizes the simulation results. Threshold estimates are presented in Table 20.

Table 19 shows that as the volatility of the upper regime decreases, the $\overline{MSE}$ and $\overline{MAE}$

for all three models decrease very significantly. The $\overline{MSE}$ ratio of TAR to RW is 0.508 in simulation 9 but 0.518 in simulation 10, showing a slight decline in the performance of TAR relative to RW. However, the S_2 measure shows increased performance of TAR and AR models with respect to RW. In simulation 11, where the underlying process is a linear AR model, it is striking to see that TAR model performs as well as an AR model. Table 20, on the other hand, shows that as the volatility of the upper regime declines, the estimates of the threshold variable become poorer. As the volatility of the upper regime gets smaller, which makes upper regime more 'similar' to the lower regime, it gets harder to identify the regime shifts. Notice that, in all three simulations the threshold value is underestimated on average and this problem becomes more severe as we move to simulation 11. Let the estimated threshold equal $\hat{r}$ while the true threshold is r. Then, in the above process, the probability of estimating a wrong regime is equal to:

$$
\begin{aligned}
\text{Prob(estimate upper|true lower)} &= \text{Prob}(\hat{r} < |Y_{t-d}| < r) \; if \; \hat{r} < r \\
\text{Prob(estimate lower|true upper)} &= \text{Prob}(r < |Y_{t-d}| < \hat{r}) \; if \; r < \hat{r}
\end{aligned}
$$

Given the evidence in Table 20, the first probability is lower for simulation 9 (relative to simulation 10) and the second probability is higher. Yet, the MSE associated with the wrong estimate of the regime is much more severe in simulation 9 relative to simulation 10 leading to much higher $\overline{MSE}$'s in table 19 for simulation 9. This result again shows that the correct estimation of the threshold becomes more important as the overall process becomes more 'nonlinear'.

In the next section we use a data set on US Dollar and Hong Kong Dollar exchange rate to illustrate the points above.

5.2.1. US/HK Exchange Rate

The data set covers two non-overlapping periods of daily exchange rates: Feb 10, 1991 to August 30, 1992 (Y_t^1) and August 20, 1993 to March 10, 1995 (Y_t^2) each giving 400 observations. Two TAR models are fitted to each data (first differences of the logs) using the first 380 observations. Table 23 presents the descriptive statistics of each data set, Table 24 results of the non-linearity tests up to 10 lags. The null hypothesis of linearity is rejected for all lags at 5 % level. Table 25 presents the parameters of the TAR models.

As it is clear from Table 23, the first data set (Y_t^1) is much more volatile than the second (Y_t^2). This can also be seen from the graph in appendix. Both of the series are distributed around zero approximately symmetrically. There are shifts in the volatility of the exchange rate without any mean shifts as in our volatility shift experiment. Table 24 shows the strong non-linearity where the null hypothesis of linearity is rejected for all lags at 5 % level. TAR models fitted in Table 25 conform to the properties of the data we present. Y_t^1 series and the fitted TAR model imply shifts in volatility, where the high volatility regime is almost three times more volatile than the low volatility regime. In Y_t^2 series, however, volatility shifts are not as pronounced as the previous one. Volatility of upper regime is about 1.4 times the volatility of the other regime.

We use the fitted models to generate data of length 750 observations and discard first 250 of them to eliminate the start up effects. Of the remaining 500 observations, first 495 are used for model fitting while last 5 observations are used for out-of-sample forecasting. We have 1000 replications in each case. Table 26, below, reports the simulation results, where simulation 12 refers to the first series and simulation 13 refers to the second.

In table 26 we see the similar patterns we observed in our volatility shift experiment. In simulation 12, where we observe large volatility shifts, threshold estimates are better than simulation 13. The $\overline{MSE}(TAR)/\overline{MSE}(RW)$ ratio in simulation 12 is 0.402 while it is 0.489 in simulation 13. This is the pattern we observed earlier in Table 19 between simulations 9 and 10. Although, better estimates of the threshold can be obtained in the first simulation, wrong prediction of the regime causes severe mispredictions. Note again the significant decline in MSE, MAE measures between the first and second simulations.

This real data example illustrates once more that, even if RW can outperform TAR or AR model in certain periods, on the average the performance of TAR and AR models are much better. Given the relevance of sign predictions, ME's under certain conditions along with other measures, the above simulations highlight the importance of knowing the loss function of the analyst who is considering the use of alternative times series tools for the problem at hand.

6. Conclusion

In this study we explore the forecast performance of TAR models under two types non linearity: (a) shifts in the mean, (b) shifts in the volatility. Using parameter specifications motivated by both real data and simulated data, we see that estimation of the delay parameter and threshold value are crucial for forecast performance of TAR models. As the nonlinearity of the process increases, the threshold estimates deteriorates. Poor threshold estimates lead to poor out of sample forecast performance. We also note that the results obtained from forecast comparisons over a couple of periods can be quite different than the results that would be obtained from comparisons by using the models repeatedly. Although

RW model can beat TAR and AR models in certain replications, on the average TAR and AR models perform much better. Conclusions derived from using single data sets over a couple of periods should therefore be treated as a single observation from a large set of possible outcomes. This study also stresses the importance of knowledge about the loss function. Depending on the summary measure used conclusions can differ.

We hope that, the results presented in this paper will encourage researchers to improve techniques used in the estimation of threshold parameters along with other model parameters. We believe that estimation of threshold parameters deserves as much attention as detecting non linearity in TAR modelling.

Footnotes

(1) To detect the threshold nonlinearity and select the threshold variable, Tsay (1989) proposed an $F-test$ based on the arranged autoregression. There are two steps in calculating the test statistic. For a given AR order p and a threshold lag d, an arranged autoregression of order p is fitted recursively to the series Y_t. Assuming that the recursion starts with the first b observations, standardized predictive residual $\widehat{e}_t$ for $t > b$ are calculated. At the second step, predictive residual $\widehat{e}_t$ are regressed on $(1, Y_{t-1}, Y_{t-2}, ..., Y_{t-p})$, and the corresponding residuals $\widehat{\varepsilon}_t$ are saved. This forms the F-statistic:

$$\widehat{F}(p,d) = \frac{\left(\sum \widehat{e}_t^2 - \sum \widehat{\varepsilon}_t^2\right)/(p+1)}{\sum \widehat{\varepsilon}_t^2/(T-d-b-p-h)} \qquad (3)$$

where $h = \max(1, p+1-d)$ and summation is over t from $b+1$ to $T-d-h+1$. Under the null hypothesis that Y_t is an $AR(p)$process, the $\widehat{F}(p,d)$ statistic is asymptotically an F distribution with degrees of freedom $p+1$ and $T-d-b-p-h$.

(2) The method used to locate the thresholds, hence the partitions of the Euclidean space, are scatterplots of various statistics versus the specified threshold variable. We use the scatterplots of the standardized predictive residuals that are obtained in step 2 of the algorithm against the specified threshold variable. Tsay (1989) proposes other alternatives as well. In an arranged autoregression framework, the TAR model consists of various model changes that occur at each threshold value r_j. A scatterplot of the standardized predictive residuals versus the threshold variable may reveal the locations of the threshold values of a TAR model since the predictive residuals will be biased at the threshold values. In a linear time series plot, however, the plot is random except for the beginning of the recursion.

(3) For an AR(1) process: $Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma_\varepsilon)$ the mean of the process is $\mu(Y_t) = \frac{\phi_0}{1-\phi_1}$ and the variance is $\sigma^2(Y_t) = \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$.

(4) The standard errors reported in parenthesis in table 5 for S_2 refer to the standard error of $S_2^i = [(MSE(TAR) - MSE(RW))/MSE(RW)]_i$ values obtained over the replications. The S_2 measure is very sensitive to the replications where performance of the TAR model is much worse than the performance of the RW. This leads to very large values of S_2^i and the large standard errors reported in the table.

Appendix

Table 1 - Ordinary and Arranged Autoregression

(a) Ordinary autoregression					(b) Arranged autoregression				
Time	Y_t	Y_{t-1}	Y_{t-2}	regime	Time	Y_t	Y_{t-1}	Y_{t-2}	regime
3	−0.41	1.21	1.31	L_2	8	0.12	−1.85	−3.08	L_1
4	0.21	−0.41	1.21	L_2	9	0.58	0.12	−1.85	L_1
5	−1.12	0.21	−0.41	L_1	7	−1.85	−3.08	−1.12	L_1
6	−3.08	−1.12	0.21	L_2	5	−1.12	0.21	−0.41	L_1
7	−1.85	−3.08	−1.12	L_1	10	1.28	0.58	0.12	L_2
8	0.12	−1.85	−3.08	L_1	6	−3.08	−1.12	0.21	L_2
9	0.58	0.12	−1.85	L_1	4	0.21	−0.41	1.21	L_2
10	1.28	0.58	0.12	L_2	3	−0.41	1.21	1.31	L_2

Shift in the Mean

Model Parameters

Table 2-Simulation 1

$r = 0$

	ϕ_0	ϕ_1	μ_ε	σ_ε	μ-regime	σ-regime
upper regime	0.01125	0.1	0	0.04975	0.0125	0.05
lower regime	−0.01125	0.1	0	0.04975	−0.0125	0.05

Table 3-Simulation 2

$r = 0$

	ϕ_0	ϕ_1	μ_ε	σ_ε	μ-regime	σ-regime
upper regime	0.0225	0.1	0	0.04975	0.025	0.05
lower regime	−0.0225	0.1	0	0.04975	−0.025	0.05

Table 4-Simulation 3

$r = 0$

	ϕ_0	ϕ_1	μ_ε	σ_ε	μ-regime	σ-regime
upper regime	0.045	0.1	0	0.04975	0.05	0.05
lower regime	−0.045	0.1	0	0.04975	−0.05	0.05

Table 5-Simulation Results (1-Step-Ahead Out-of-Sample)

	Simulation1	Simulation2	Simulation3
$\overline{MSE}$(TAR)	0.0024872(0.00155)	0.0026175(0.00158)	0.0028197(0.00194)
$\overline{MSE}$(AR)	0.0024801(0.00155)	0.0026411(0.00162)	0.0030107(0.00196)
$\overline{MSE}$(RW)	0.0038749(0.00282)	0.0037138(0.00267)	0.0036834(0.00270)
S_2(TAR/AR)	0.597(8.4)	-0.084(19.38)	-2.633(56.4)
S_2(TAR/RW)	-22.33(60.3)	-14.74(83.8)	-11.874(66.5)
S_2(AR/RW)	-22.51(57.9)	-13.15(88.9)	-7.259(51.6)
$\overline{MAE}\,(TAR)$	0.039217(0.0132)	0.040992(0.0132)	0.042161(0.0153)
$\overline{MAE}\,(AR)$	0.039841(0.0132)	0.041378(0.0133)	0.043748(0.0152)
$\overline{MAE}\,(RW)$	0.049545(0.0186)	0.048258(0.0179)	0.048067(0.0184)
$\overline{ME}\,(TAR)$	-1.51e-5(0.022)	1.624e-3(0.022)	-4.5e-4(0.024)
$\overline{ME}\,(AR)$	0.469e-5(0.022)	1.505e-3(0.022)	-8.4e-4(0.025)
$\overline{ME}\,(RW)$	-34.9e-5(0.014)	2.85e-4(0.015)	-2.9e-4(0.018)
$\overline{SP}\,(TAR)$	67.2(17.8)	62.5(18.7)	61.2(18.7)
$\overline{SP}\,(AR)$	67.2(17.9)	61.02(17.9)	55.8(16.8)
$\overline{SP}\,(RW)$	0(0)	0(0)	0(0)

Table 6-Simulation Results (3-Step-Ahead Out-of-Sample)

	Simulation1	*Simulation2*	*Simulation3*
$\overline{MSE}(TAR)$	2.668e-3(2.17e-3)	3.3e-3(2.5e-3)	4.5e-3(4.02e-3)
$\overline{MSE}(AR)$	2.641e-3(2.17e-3)	3.2e-3(2.4e-3)	4.46e-3(3.56e-3)
$\overline{MSE}(RW)$	5.349e-3(4.3e-3)	6.01e-3(5.05e-3)	6.44e-3(6.30e-3)
$\overline{MAE}(TAR)$	4.12e-2(1.8e-2)	4.65e-2(1.98e-2)	5.38e-2(2.62e-2)
$\overline{MAE}(AR)$	4.10e-2(1.81e-2)	4.57e-2(1.93e-2)	5.41e-2(2.39e-2)
$\overline{MAE}(RW)$	5.8e-2(2.53e-2)	6.2e-2(2.82e-2)	6.32e-2(3.30e-2)
$\overline{ME}(TAR)$	3.27e-4(3.5e-2)	3.9e-3(4.2e-2)	2.82e-3(5.21e-2)
$\overline{ME}(AR)$	1.007e-5(3.58e-2)	3.0e-3(4.2e-2)	-1.9e-3(5.38e-2)
$\overline{ME}(RW)$	-8.8e-4(4.7e-2)	1.68e-3(5.4e-2)	-2.8e-5(5.9e-2)
$\overline{SP}(TAR)$	74.3(25.59)	72.6(26.1)	64.2(29.1)
$\overline{SP}(AR)$	74.2(25.7)	73.3(26.8)	63.8(29.3)
$\overline{SP}(RW)$	0(0)	0(0)	0(0)

Table 7-Simulation Results (5-Step-Ahead Out-of-Sample)

	*Simulation*1	*Simulation*2	*Simulation*3
$\overline{MSE}(TAR)$	2.55e-3(3.6e-3)	3.47e-3(4.45e-3)	5.39e-3(7.3e-3)
$\overline{MSE}(AR)$	2.52e-3(3.6e-3)	3.3e-3(4.18e-3)	4.91e-3(6.0e-3)
$\overline{MSE}(RW)$	5.25e-3(7.5e-3)	6.3e-3(8.39e-3)	8.29e-3(1.3e-2)
$\overline{MAE}(TAR)$	4.02e-2(3.06e-2)	4.8e-2(3.42e-2)	5.85e-2(4.4e-2)
$\overline{MAE}(AR)$	3.98e-2(3.05e-2)	4.7e-2(3.3e-2)	5.76e-2(3.99e-2)
$\overline{MAE}(RW)$	5.83e-2(4.3e-2)	6.4e-2(4.7e-2)	7.10e-2(5.70e-2)
$\overline{ME}(TAR)$	1.02e-3(5.05e-2)	3.9e-3(5.8e-2)	6.0e-3(7.3e-2)
$\overline{ME}(AR)$	4.9e-4(5.02e-2)	3.0e-3(5.7e-2)	-8.9e-4(7.0e-2)
$\overline{ME}(RW)$	-1.7e-3(7.25e-2)	1.43e-3(7.9e-2)	1.5e-3(9.1e-2)
$\overline{SP}(TAR)$	75.4(43.08)	73.6(44.1)	68.8(46.4)
$\overline{SP}(AR)$	76(42.7)	73.9(43.9)	70.0(45.8)
$\overline{SP}(RW)$	0(0)	0(0)	0(0)

Shift in the Mean - Threshold estimation

Table 8-Threshold Estimates (True r=0)

	Simulation 1	Simulation 2	Simulation3
min	-0.069	-0.074	-0.093
mean	-0.016	-0.017	-0.024
max	0.114	0.127	0.143
std	0.0334	0.0333	0.0365
median	-0.025	-0.022	-0.023
skewness	1.108	0.927	0.449

Table 9-Effect of Threshold Estimation

	$\overline{MSE}\,(TAR)$	$S_2\,(TAR)$
Sim1	0.00248	-22.33
Sim4	0.00246	-24.44
S4/S1	0.992	1.09
Sim2	0.00261	-14.75
Sim5	0.00245	-18.68
S5/S2	0.938	1.26
Sim3	0.00282	-11.87
Sim6	0.00240	-19.75
S6/S3	0.851	1.66

UK/DM Exchange Rate Data
Shift in the Mean

Table 10-UK/DM

	min	mean	max	std	median	skewness
Y_t	-0.0142	$7.4e-5$	0.0258	$3.29e-3$	$1.7e-5$	$1.45e-8$
Y_t^1	-0.0142	$9.53e-5$	0.0258	$3.33e-3$	$3.4e-5$	$1.39e-8$
Y_t^2	-0.0142	$1.46e-5$	0.0258	$3.25e-3$	0	$1.47e-8$

Table 11-Test of Non-linearity

	Y_t			Y_t^1			Y_t^2		
d	F-stat	n-df	d-df	F-stat	n-df	d-df	F-stat	n-df	d-df
1	92.55	2	447	89.91	2	430	68.75	2	430
2	10.31	2	446	10.18	2	429	9.803	2	429
3	50.22	2	445	46.73	2	428	47.59	2	428
4	3.323	2	444	2.972	2	427	6.970	2	427
5	3.608	2	443	3.048	2	426	2.566	2	426
6	13.75	2	442	11.57	2	425	20.29	2	425
7	33.13	2	441	33.33	2	424	45.19	2	424
8	82.73	2	440	79.21	2	423	78.58	2	423
9	2.841	2	439	3.256	2	422	4.058	2	422
10	4.937	2	438	4.191	2	421	6.250	2	4219

Table 12-model estimates $\left(\widehat{r}=1.03\text{e-}4,\ \widehat{d}=1\right)$

	Y_t^1	Y_t^1	Y_t^2	Y_t^2
	lower regime	upper regime	lower regime	upper regime
$\widehat{\phi}_0^i$	0	0.00027	0	0.00017
$\widehat{\phi}_1^i$	0.0542	-0.10978	0.0408	-0.1466
$\mu\left(Y_t^i\right)$	0	2.43e-4	0	1.48e-4
$\sigma\left(Y_t^i\right)$	0.00318	0.0035	0.00303	0.0035
μ_ε^i	0	0	0	0
σ_ε^i	0.00317	0.003446	0.00300	0.00346

Table 13-Simulation Results (1-Step-Ahead, Out-of-sample)

	Simulation 7	Simulation 8
$\overline{MSE}(TAR)$	1.1407e-5(6.98e-6)	1.0690e-5(6.67e-6)
$\overline{MSE}(AR)$	1.1409e-5(6.99e-6)	1.0639e-5(6.67e-6)
$\overline{MSE}(RW)$	2.2462e-5(1.67e-6)	2.1866e-5(1.68e-5)
$\overline{MAE}\,(TAR)$	0.002714(0.0009)	0.002597(0.00087)
$\overline{MAE}\,(AR)$	0.002714(0.0009)	0.002588(0.00086)
$\overline{MAE}\,(RW)$	0.003802(0.00150)	0.003707(0.00151)
$\overline{ME}\,(TAR)$	-6.20e-5(0.00152)	5.45e-5(0.00146)
$\overline{ME}\,(AR)$	-6.29e-5(0.00151)	5.58e-5(0.00146)
$\overline{ME}\,(RW)$	2.50e-5(0.00091)	3.36e-5(0.00089)
$\overline{SP}\,(TAR)$	0.7456(0.175)	0.7356(0.178)
$\overline{SP}\,(AR)$	0.7422(0.176)	0.7374(0.176)
$\overline{SP}\,(RW)$	0(0)	0(0)
$\overline{\hat{r}}$	-0.00077	-0.00101
$\sigma_{\hat{r}}$	0.00216	0.00208
$\min(\hat{r})$	-0.0044	-0.0044
$\max(\hat{r})$	0.0072	0.0076

Table 14-Simulation Results 3-Step-Ahead, Out-of-sample)

	Simulation 7	Simulation 8
$\overline{MSE}(TAR)$	1.12e-5(8.56e-6)	1.07e-5(9.15e-6)
$\overline{MSE}(AR)$	1.12e-5(8.51e-6)	1.07e-5(9.14e-6)
$\overline{MSE}(RW)$	2.30e-5(1.81e-5)	2.07e-5(1.67e-5)
$\overline{MAE}(TAR)$	2.7e-3(1.11e-3)	2.6e-3(1.16e-3)
$\overline{MAE}(AR)$	2.7e-3(1.11e-3)	2.6e-3(1.16e-3)
$\overline{MAE}(RW)$	3.8e-3(1.65e-3)	3.64e-3(1.60e-3)
$\overline{ME}(TAR)$	-1.5e-4(1.9e-3)	-5.17e-5(1.81e-3)
$\overline{ME}(AR)$	-5.53e-5(1.95e-3)	5.83e-5(1.80e-3)
$\overline{ME}(RW)$	3.16e-5(2.73e-3)	5.92e-5(2.54e-3)
$\overline{SP}(TAR)$	75.5(25.2)	75.5(25.2)
$\overline{SP}(AR)$	75.3(25.2)	75.3(24.8)
$\overline{SP}(RW)$	0(0)	0(0)

Table 15-Simulation Results 5-Step-Ahead, Out-of-sample)

	Simulation 7	Simulation 8
$\overline{MSE}(TAR)$	1.09e-5(1.51e-5)	1.057e-5(1.42e-5)
$\overline{MSE}(AR)$	1.09e-5(1.50e-5)	1.056e-5(1.42e-5)
$\overline{MSE}(RW)$	2.062e-5(2.93e-5)	2.02e-5(2.69e-5)
$\overline{MAE}\,(TAR)$	2.61e-3(2.01e-3)	2.60e-3(1.94e-3)
$\overline{MAE}\,(AR)$	2.61e-3(2.01e-3)	2.60e-3(1.94e-3)
$\overline{MAE}\,(RW)$	3.64e-3(2.7e-3)	3.62e-3(2.66e-3)
$\overline{ME}\,(TAR)$	-9.6e-5(3.3e-3)	-3.01e-5(3.25e-3)
$\overline{ME}\,(AR)$	4.26e-6(3.3e-3)	8.09e-5(3.25e-3)
$\overline{ME}\,(RW)$	1.25e-4(4.5e-3)	1.68e-4(4.5e-3)
$\overline{SP}\,(TAR)$	73.1(44.36)	74.0(43.8)
$\overline{SP}\,(AR)$	72.6(44.6)	73.5(44.1)
$\overline{SP}\,(RW)$	0(0)	0(0)

Shift in the Volatility

Model Parameters

Table 16-Simulation 9

$r = 0.00085, d = 1$

	ϕ_0	ϕ_1	μ_ε	σ_ε	μ-regime	σ-regime
upper regime	0	0.01	0	0.003	0	3e-3
lower regime	0	0.01	0	0.0005	0	5e-4

Table 17-Simulation 10

$r = 0.00085, d = 1$

	ϕ_0	ϕ_1	μ_ε	σ_ε	μ-regime	σ-regime
upper regime	0	0.01	0	0.0015	0	1.5e-3
lower regime	0	0.01	0	0.0005	0	5e-4

Table 18-Simulation 11

$r = 0.00085, d = 1$

	ϕ_0	ϕ_1	μ_ε	σ_ε	μ-regime	σ-regime
upper regime	0	0.01	0	0.0005	0	5e-4
lower regime	0	0.01	0	0.0005	0	5e-4

Shift in the Volatility - Simulation Results

Table 19-Simulation Results (1-Step-Ahead Out-of-Sample)

	Simulation 9	Simulation10	Simulation11
$\overline{MSE}(TAR)$	2.87e-6(4.78e-6)	6.33e-7(9.89e-7)	2.447e-7(1.51e-7)
$\overline{MSE}(AR)$	2.85e-6(4.76e-6)	6.27e-7(9.78e-7)	2.439e-7(1.51e-7)
$\overline{MSE}(RW)$	5.64e-6(1.038e-5)	1.22e-6(1.88e-6)	4.93e-7(3.76e-7)
$S_2(TAR/AR)$	1.77(19)	1.32(10)	0.66(6)
$S_2(TAR/RW)$	-26.37(94)	-32.3(63)	-35.6(65)
$S_2(AR/RW)$	-26.10(97)	-32.9(61)	-36.1(67)
$\overline{MAE}\,(TAR)$	9.93e-4(9.6e-4)	5.52e-4(3.7e-4)	3.95e-4(1.3e-4)
$\overline{MAE}\,(AR)$	9.91e-4(9.58e-4)	5.50e-4(3.67e-4)	3.94e-4(1.3e-4)
$\overline{MAE}\,(RW)$	13.5e-4(13.48e-4)	7.52e-4(5.1e-4)	5.61e-4(2.2e-4)
$\overline{ME}\,(TAR)$	7.057e-6(7.6e-4)	1.447e-5(3.65e-4)	1.21e-6(2.24e-4)
$\overline{ME}\,(AR)$	1.356e-5(7.5e-4)	-1.474e-5(3.61e-4)	1.63e-6(2.22e-4)
$\overline{ME}\,(RW)$	1.52e-6(4.7e-4)	-8.49e-6(2.3e-4)	2.82e-6(1.4e-4)
$\overline{SP}\,(TAR)$	73.5(18.2)	75.5(18.1)	75.1(17.5)
$\overline{SP}\,(AR)$	73.3(18.1)	75.4(18)	75.2(17.6)
$\overline{SP}\,(RW)$	0(0)	0(0)	0(0)

Shift in the Volatility - Threshold Estimation

Table 20-Threshold Estimates (True r=0.00085)

$\widehat{d}$=1

	Simulation 9	Simulation 10	Simulation 11
min	0.00012	0.00009	0.00009
mean	0.00063	0.00057	0.00045
max	0.00096	0.00098	0.00131
std	0.00023	0.00026	0.00023
median	0.00066	0.00054	0.00040
skewness	-8.4e-11	-7.8e-12	-3.77e-11

Shift in Volatility 3-step and 5-step Ahead Forecast Results

Table 21-Simulation Results 3-Step-Ahead Out-of-Sample)

	Simulation 9	Simulation10	Simulation11
$\overline{MSE}(TAR)$	2.78e-6(5.55e-6)	6.077e-7(1.092e-6)	2.37e-7(1.95e-7)
$\overline{MSE}(AR)$	2.79e-6(5.57e-6)	6.072e-7(1.091e-6)	2.36e-7(1.94e-7)
$\overline{MSE}(RW)$	5.67–E6(1.01e-5)	1.24e-6(2.13e-6)	4.74e-7(3.75e-7)
$\overline{MAE}(TAR)$	9.79e-4(1.07e-3)	5.44e-4(4.14e-4)	3.88e-4(1.7e-4)
$\overline{MAE}(AR)$	9.80e-4(1.07e-3)	5.43e-4(4.14e-4)	3.87e-4(1.7e-4)
$\overline{MAE}(RW)$	1.49e-3(1.44e-3)	7.84e-4(5.85e-4)	5.52e-4(2.33e-4)
$\overline{ME}(TAR)$	3.49e-5(9.93e-4)	-1.70e-5(4.63e-4)	2.07e-6(2.79e-4)
$\overline{ME}(AR)$	3.56e-5(9.96e-4)	-1.57e-5(4.62e-4)	2.19e-6(2.79e-4)
$\overline{ME}(RW)$	6.14e-5(1.37e-3)	-1.42e-5(6.84e-4)	6.27e-6(3.9e-4)
$\overline{SP}(TAR)$	75(27.2)	73.73(25.9)	74.93(24.5)
$\overline{SP}(AR)$	74.4(26.9)	73.76(25.65)	75.1(24.4)
$\overline{SP}(RW)$	0(0)	0(0)	0(0)

Table 22-Simulation Results 5-Step-Ahead Out-of-Sample)

	Simulation 9	Simulation10	Simulation11
$\overline{MSE}(TAR)$	2.81e-6(7.62e-6)	5.45e-7(1.37e-6)	2.20e-7(2.98e-7)
$\overline{MSE}(AR)$	2.82e-6(7.64e-6)	5.44e-7(1.35e-6)	2.20e-7(2.99e-7)
$\overline{MSE}(RW)$	5.55e-6(1.29e-5)	1.31e-6(2.89e-6)	4.97e-7(6.72e-7)
$\overline{MAE}(TAR)$	9.83e-4(1.36e-3)	5.25e-4(5.20e-4)	3.75e-4(2.81e-4)
$\overline{MAE}(AR)$	9.85e-41(1.36e-3)	5.25e-4(5.18e-4)	3.74e-4(2.82e-4)
$\overline{MAE}(RW)$	1.52e-3(1.79e-3)	8.11e-4(8.09e-4)	5.57e-4(4.18e-4)
$\overline{ME}(TAR)$	-4.50e-5(1.68e-3)	-2.23e-5(7.38e-4)	1.02e-6(4.69e-4)
$\overline{ME}(AR)$	-4.14e-5(1.68e-3)	-2.10e-5(7.38e-4)	1.00e-6(4.69e-4)
$\overline{ME}(RW)$	7.61e-6(2.35e-3)	-4.24e-5(1.14e-3)	1.41e-5(7.05e-4)
$\overline{SP}(TAR)$	76.2(42.6)	76.1(42.6)	78.7(40.96)
$\overline{SP}(AR)$	76.3(42.5)	75.7(42.9)	78.4(41.17)
$\overline{SP}(RW)$	0(0)	0(0)	0(0)

US/HK Exchange Rate Data
Shift in the Volatility

Table 23-US/HK

	min	mean	max	std	median	skewness
Y_t^1	-0.0048	2.27e-5	0.0041	0.0007	0	-1.57e-10
Y_t^2	-0.00209	-3.48e-6	0.0017	0.00028	0	-1.902e-11

Table 24-Test of Non-linearity

	Y_t^1			Y_t^2		
d	F-stat	n-df	d-df	F-stat	n-df	d-df
1	94.72	2	338	15.93	2	338
2	10.74	2	337	7.137	2	337
3	5.119	2	336	13.16	2	336
4	12.89	2	335	10.97	2	335
5	11.12	2	334	13.90	2	334
6	10.13	2	333	63.77	2	333
7	1.000	2	332	11.73	2	332
8	1.218	2	331	14.90	2	331
9	31.56	2	330	56.33	2	330
10	44.33	2	329	1.74	2	329

Table 25-model estimates

$\hat{r} = 0.0009$, $\hat{d} = 1$ $\hat{r} = 0.00059$, $\hat{d} = 1$

	Y_t^1	Y_t^1	Y_t^2	Y_t^2
	lower regime	upper regime	lower regime	upper regime
$\widehat{\phi}_0^i$	0	0	0	0
$\widehat{\phi}_1^i$	0.02939	-0.29739	-0.04018	-0.049578
$\mu(Y_t^i)$	0	0	0	0
$\sigma(Y_t^i)$	0.000578	0.001462	0.000278	0.000385
μ_ε^i	0	0	0	0
σ_ε^i	0.000578	0.001328	0.000278	0.0003721

Table 26-Simulation Results (1-Step-Ahead, Out-of-sample)

	Simulation 12	Simulation 13
$\overline{MSE}(TAR)$	6.625e-7(8.07e-7)	7.86e-8(5.48e-8)
$\overline{MSE}(AR)$	6.667e-7(8.07e-7)	7.85e-8(5.51e-8)
$\overline{MSE}(RW)$	16.48e-7(24.9e-7)	16.05e-8(12.7e-8)
$\overline{MAE}(TAR)$	6.01e-4(3.2e-4)	2.22e-4(8.05e-5)
$\overline{MAE}(AR)$	6.04e-4(3.3e-4)	2.22e-4(8.04e-5)
$\overline{MAE}(RW)$	8.95e-4(6.0e-4)	3.19e-4(1.28e-4)
$\overline{ME}(TAR)$	1.404e-5(3.6e-4)	7.71e-6(1.25e-4)
$\overline{ME}(AR)$	1.378e-5(3.6e-4)	7.58e-6(1.26e-4)
$\overline{ME}(RW)$	1.05e-5(2.2e-4)	-2.06e-6(8.02e-5)
$\overline{SP}(TAR)$	75.7(17.1)	75.4(17.5)
$\overline{SP}(AR)$	75.5(17.2)	75.54(17.5)
$\overline{SP}(RW)$	0(0)	0(0)
$\overline{\hat{r}}$	0.00068	0.000286
$\sigma_{\hat{r}}$	0.000275	0.000164
$\min(\hat{r})$	0.000114	4.96e-5
$\max(\hat{r})$	0.0015	0.00075

Table 27-Simulation Results (3-Step-Ahead, Out-of-sample)

	Simulation 12	Simulation 13
$\overline{MSE}(TAR)$	6.89e-7(1.08e-6)	7.74e-8(6.97e-4)
$\overline{MSE}(AR)$	6.87e-7(1.07e-6)	7.71e-8(6.92e-8)
$\overline{MSE}(RW)$	1.40e-6(2.01e-6)	1.61e-7(1.40e-7)
$\overline{MAE}(TAR)$	6.08e-4(4.2e-4)	2.21e-4(1.02e-4)
$\overline{MAE}(AR)$	6.06e-4(4.18e-4)	2.20e-4(1.02e-4)
$\overline{MAE}(RW)$	8.81e-4(5.63e-4)	3.2e-4(1.49e-4)
$\overline{ME}(TAR)$	1.89e-5(4.23e-4)	9.79e-6(1.64e-4)
$\overline{ME}(AR)$	1.99e-5(4.20e-4)	9.41e-6(1.63e-4)
$\overline{ME}(RW)$	3.28e-5(6.05e-4)	3.59e-6(2.28e-4)
$\overline{SP}(TAR)$	74.86(25.95)	76.16(24.54)
$\overline{SP}(AR)$	75.26(25.90)	76.23(24.79)
$\overline{SP}(RW)$	0(0)	0(0)

Table 28-Simulation Results (5-Step-Ahead, Out-of-sample)

	Simulation 12	Simulation 13
$\overline{MSE}(TAR)$	6.67e-7(1.51e-6)	7.92e-8(1.07e-7)
$\overline{MSE}(AR)$	6.63e-7(1.49e-6)	7.86e-8(1.06e-7)
$\overline{MSE}(RW)$	1.29e-6(2.41e-6)	1.61e-7(2.09e-7)
$\overline{MAE}\,(TAR)$	5.87e-4(5.68e-4)	2.26e-4(1.67e-4)
$\overline{MAE}\,(AR)$	5.84e-4(5.67e-4)	2.27e-4(1.67e-4)
$\overline{MAE}\,(RW)$	8.51e-4(7.51e-4)	3.26e-4(2.33e-4)
$\overline{ME}\,(TAR)$	2.24e-6(8.17e-4)	-6.21e-7(2.81e-4)
$\overline{ME}\,(AR)$	3.11e-6(8.14e-4)	-1.17e-6(2.81e-4)
$\overline{ME}\,(RW)$	5.25e-5(1.13e-3)	-1.03e-5(4.01e-4)
$\overline{SP}\,(TAR)$	76.7(42.29)	74.7(43.49)
$\overline{SP}\,(AR)$	77.1(42.04)	74.4(43.66)
$\overline{SP}\,(RW)$	0(0)	0(0)

Tests of Significance

Table 29 and 30, below, gives the results of two tests : Sign test and Wilcoxon's Signed-Rank Test. For Sign test the null hypothesis is that the difference between the two series (e.g. difference between MSE's of TAR and RW model along 1000 replications) has a median of zero. If the series is symmetrically distributed, then in the earlier null median and mean are equal. (In first 9 columns symmetry is satisfied). If we let d_t refer to the difference series obtained by subtracting two series (e.g. MSE(TAR)-MSE(RW)), then for a sample size of T observations, the test statistic is simply:

$$K_2 = \sum_{t=1}^{T} \frac{K_{2a} - .5T}{\sqrt{.25T}} \sim N(0,1)$$

where $K_{2a} = \sum_{t=1}^{T} I_+ (d_t)$ and

$$\begin{aligned} I_+ (d_t) &= 1 \, if \; d_t > 0 \\ &= 0 \; o.w. \end{aligned}$$

Wlicoxon's Signed Rank Test (with same null hypothesis of equal median) requires symmetry of d_t series (but can be more powerful than Sign test in that case). The test statistic

$$K_3 = \sum_{t=1}^{T} I_+ (d_t) \, rank(|d_t|)$$

is the sum of the ranks of the absolute values of the positive observations. Its studentized version is asymptotically standard normal:

$$K_{3a} = \frac{K_3 - \frac{T(T+1)}{4}}{\sqrt{\frac{T(T+1)(2T+1)}{24}}} \sim N(0,1).$$

In Tables 29 and 30, rows refer to the simulations reported in the text.(Table 30 is the continuation of the Table 29). Each block of three columns refer to MSE, MAE, ME, SP AND S_2 measures respectively. For example, the cell referring to first column and first row is the matrix: $\boxed{\begin{array}{c} \text{1-1-1} \\ \hline \text{0-1-1} \end{array}}$. The first row of the matrix refers to the sign test for 1-step, 3-step and 5-step ahead out-of-sample forecasts of TAR and AR models in simulation 1. (1) indicates that H_0 is rejected at 5 % level whereas, (0) indicates that H_0 can not be rejected at 5 % level. Second row refers to the Wilcoxon's Signed Rank test. For columns 10-15 we don't report the result of Wilcoxon test because the requirement of symmetry is not satisfied for these samples.

Table 29-Tests of Significance

	MSE			MAE			ME	
	T-AR	T-R	AR-R	T-AR	T-R	AR-R	T-AR	T-R
Sim 1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-0-0	0-0-1
	0-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	0-1-1	0-0-0
Sim 2	0-1-1	1-1-1	1-1-1	0-1-1	1-1-1	1-1-1	1-1-1	0-1-1
	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	0-1-1	1-1-1
Sim 3	1-0-0	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1	0-1-1	0-0-1
	1-0-1	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1	0-1-1	0-1-1
Sim 7	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	0-0-0
	0-0-0	1-1-1	1-1-1	0-0-0	1-1-1	1-1-1	0-1-1	0-1-1
Sim 8	0-1-1	1-1-1	1-1-1	0-1-1	1-1-1	1-1-1	1-1-1	0-0-0
	1-0-0	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1	0-1-1	0-0-0
Sim 9	0-1-0	1-1-1	1-1-1	0-0-0	1-1-1	1-1-1	0-0-0	0-0-0
	0-0-0	1-1-1	1-1-1	0-0-0	1-1-1	1-1-1	1-0-0	0-0-0
Sim 10	1-0-0	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1	0-0-0	0-0-0
	1-0-0	1-1-1	1-1-1	1-0-1	1-1-0	1-1-1	0-0-1	0-0-1
Sim 11	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	0-0-0
	1-0-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	0-1-0	0-1-0
Sim 12	1-0-0	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1	0-0-0	0-0-0
	1-0-1	1-1-1	1-1-1	1-0-1	1-1-1	1-1-1	0-0-0	0-0-1
Sim 13	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	0-0-0
	1-1-1	1-1-1	1-1-1	0-1-1	1-1-1	1-1-1	0-0-1	1-0-0

Table 30-Tests of Significance

	ME	SP			S_2		
	A-R	T-AR	T-R	AR-R	T-AR	T-R	A-R
Sim 1	0-0-0	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 2	0-1-0	1-1-1	1-1-1	1-1-1	0-1-1	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 3	0-0-0	1-1-1	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 7	0-0-0	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 8	0-1-0	1-1-1	1-1-1	1-1-1	0-1-1	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 9	1-0-0	1-1-1	1-1-1	1-1-1	0-1-0	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 10	0-0-0	1-1-1	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 11	0-0-0	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 12	0-0-0	1-1-1	1-1-1	1-1-1	1-0-0	1-1-1	1-1-1
	0-0-0	-	-	-	-	-	-
Sim 13	0-0-0	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1
	1-0-0	-	-	-	-	-	-

The following table (continued in table 32) gives the test results of Sign test and Wilcoxon test described earlier. For a given measure (i.e. row) and a column the null hypothesis of no difference between the medians of the two series obtained in two different simulations is rejected if the entry is 1 and null can not be rejected if entry is 0 at 5 % level (median is equal to mean for the first 9 rows due to symettry.) See the note before table 9 for details. Wilcoxon test results are not reported for rows 10 to 15 since the requirement of symettry is not satisfied for these measures.

Table 31-Tests of Significance

	S1-S2	S2-3	S7-S8	S9-S10	S10-S11	12-13
MSE-T	1-1-1	0-1-1	1-1-0	1-1-1	1-1-1	1-1-1
	1-1-1	0-1-1	1-1-0	1-1-1	1-1-1	1-1-1
MSE-A	1-1-1	1-1-1	1-1-0	1-1-1	1-1-1	1-1-1
	1-1-1	1-1-1	1-1-0	1-1-1	1-1-1	1-1-1
MSE-R	0-1-0	0-0-0	0-1-0	1-1-1	1-1-1	1-1-1
	0-1-1	0-0-1	0-1-0	1-1-1	1-1-1	1-1-1
MAE-T	1-1-1	0-1-1	1-1-0	1-1-1	1-1-1	1-1-1
	1-1-1	0-1-1	1-1-0	1-1-1	1-1-1	1-1-1
MAE-A	1-1-1	1-1-1	1-1-0	1-1-1	1-1-1	1-1-1
	1-1-1	1-1-1	1-1-0	1-1-1	1-1-1	1-1-1
MAE-R	0-1-0	0-0-0	0-0-0	1-1-1	1-1-1	1-1-1
	0-1-1	0-0-1	0-1-0	1-1-1	1-1-1	1-1-1
ME-T	0-0-0	1-0-0	0-0-0	0-0-0	0-0-0	0-0-0
	0-0-0	1-0-0	0-0-0	0-0-0	0-0-0	0-0-0

Table 32-Tests of Significance

	S1-S2	S2-3	S7-S8	S9-S10	S10-S11	12-13
ME-A	0-0-0	1-0-1	0-0-0	0-0-0	0-0-0	0-0-0
	0-0-0	1-1-0	0-0-0	0-0-0	0-0-0	0-0-0
ME-R	0-0-0	0-0-0	0-0-0	0-0-0	0-0-0	0-0-0
	0-0-0	0-0-0	0-0-0	1-0-0	0-0-0	1-0-1
SP-T	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1
	-	-	-	-	-	-
SP-A	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1
	-	-	-	-	-	-
SP-R	0-0-0	0-0-0	0-0-0	0-0-0	0-0-0	0-0-0
	-	-	-	-	-	-
S_2-T	1-0-0	1-0-0	0-0-0	0-0-0	0-0-0	1-0-0
	-	-	-	-	-	-
S_2-A	1-0-0	0-1-1	0-0-0	0-1-0	0-0-0	0-0-0
	-	-	-	-	-	-
S_2-R	1-0-0	1-1-1	0-0-0	0-1-0	0-0-0	1-0-0
	-	-	-	-	-	-

References

[1] Aydemir, A. B. (1996), Threshold Models in Time Series Analysis with an Application to U.S. Interest Rates, *mimeo*, U.W.O.

[2] Boothe, P. and D. Glassman (1987), The Statistical Distribution of Exchange Rates, *J of International Economics*, 22, 297-319

[3] Byers, J. D. and D. A. Peel (1995), Forecasting Industrial Production Using Non-Linear Methods, *J of Forecasting,* Vol 14, 325-336

[4] Cao, C. Q. and R. S. Tsay (1992), Non-Linear Time Series Analysis of Stock Volatilities, *J of Applied Econometrics*, Vol 7, 165-185

[5] Chan, W. S. (1990), On Tests for Non-Linearity in Hong-Kong Stock Returns, Hong Kong, *J of Business Management*, VIII, 1-11

[6] Chan, W.S. and S. H. Cheung (1994), On Robust Estimation of Threshold Autoregression, *J of Forecasting*, Vol 13, 37-49

[7] Clements, M. P. and J. Smith (1996), A Monte Carlo Study of the Forecasting Performance of Empirical SETAR Model, *Warwick Economic Research Papers*, no. 464, Department of Economics, The University of Warwick (forthcoming, *J of Applied Econometrics).*

[8] Chinn. M. D. (1991), Some Linear and Non-Linear Thoughts on Exchange Rates, *J of International Money and Finance*, 10, 214-230

[9] Chappell, D., J. Padmore, P. Mistry and C. Ellis (1996), A Threshold Model for the French Franc/Deutcshemark Exchange Rate, *J of Forecasting,* Vol 15, No3

[10] Cootner, P. ed. (1964), *The Random Character of Stock Market Prices*, MIT Press, Cambridge, MA

[11] Dacco, R. and S. Satchell (1995), Why Do Regime Switching Models Forecast So Badly ?, *Birkbeck College, U of London, Discussion Paper in Financial Economics,* FE-7/95

[12] De Gooijer J. G. and K. Kumar (1992), Some Recent Developments in Non-Linear Time Series Modelling, Testing and Forecasting, *International J of Forecasting*, 8, 135-156

[13] Diebold F. X. and J. A. Nason (1990), Nonparametric Exchange Rate Prediction, *J of International Economics*, 28, 315-332

[14] Diebold, F. X. and M. Nerlove (1989), The Dynamics of Exchange Rate Volatility: A Multivariate Latent-factor ARCH Model, *J of Applied Econometrics*, 4, 1-22

[15] Domowitz, I. and C. S. Hakkio (1985), Conditional Variance and the Risk Premium in the Foreign Exchange Rate Market, *J of International Economics,* 19, 47-66

[16] Engel, C. (1994), Can the Markov Switching Model Forecast Exchange Rates?, *J of International Economics*, 36, 151-165

[17] Engel, C. M. and J. D. Hamilton (1990), Long Swings in the Dollar: Are They in the Data Set and Do Markets Know It?, *American Economic Review*, 80, 689-71

[18] Edison, H. J. (1991), Forecast Performance of Exchange Rate Models Revisited, *Applied Economics*, 23, 187-196

[19] Engel, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of Variance of U.K. Inflation, *Econometrica*, 50, 987-1008

[20] Engel, R. F., D. M. Lillien and R. P. Robbins (1987), Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model, *Econometrica*, 55, 391-408

[21] Fama, E. F. (1965), The Behavior of Stock Market Prices, J of Business, 38, 34-105

[22] Gallant, A. R., D. Hsieh and G. Tauchen (1988), On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate: 1974-1983, *manuscript*, Graduate School of Business, U of Chicago, Chicago, IL

[23] Goering, G. E. and M. K. Pippenger (1994), A Note Regarding ARCH and Threshold Processes: Results from a Monte Carlo Study, *Applied Economics Letters,* 1, 210-213

[24] Hsieh, D. A. (1989), Testing for Non-Linear Dependence in Foreign Exchange Rates: 1974-1983, *J of Business*, 62, 339-368

[25] Kuan, C. M. and T. Lin (1995), Forecasting Exchange Rates Using Feedward and Recurrent Networks, *J of Applied Econometrics*, Vol 10, 347-364

[26] Le Baron, B. (1992), Forecast Improvements Using a Volatility Index, *J of Applied Econometrics*, Vol 7, 137-149

[27] Leitch, G. and J. E. Tanner (1991), Economic Forecast Evaluation: Profits versus the Conventional Error Measures, *American Economic Review*, 81, 580-590

[28] Levich, R. (1981), How to Compare Chance with Forecasting Expertice, *Euromoney*, August, 61-78

[29] Meese, R. A. and K. Rogoff (1983a), Empirical Exchange Rate Models of the Seventies: Do They Fit out of Sample?, *J of International Economics*, 14, 3-24

[30] Meese, R. A. and K. Rogoff (1983b), The out of Sample Failure of Empirical Exchange Rate Models: sampling Error or Misspecification ?, in J. Frenkel ed., *Exchange Rates and International Economics*, U of Chicago Press, Chicago, IL

[31] Merton, R. C. (1981), On the Market Timing and Investment Performance, I: An Equilibrium Theory of Value for Market Forecasts, *J of Business*, 54, 363-406

[32] Ozaki, T. and H. Tong (1975), On Fitting of Non-Stationary Autoregressive Models in Time Series Analysis, in *Proceedings of 8th Hawai International Conference on System Sciences,* 225-226, North Hollywood, Western Periodical

[33] Peel, D. A. and A. E. H. Speight (1994), Testing for Non-Linear Dependence in Inter-War Exchange Rates, *Weltwirtschaftlisches Archiv,* 130 (2), 391-417

[34] Priestly, M. B. (1965), Evolutionary Spectra and Non-Stationary Processes, *J of Royal Statistical Society,* B, 27, 204-237

[35] Priestly, M. B. and H. Tong (1973), On the Analysis of Bivariate Non-Stationary Processes, *J of Royal Statistical Society,* B, 35, 153-166, 179-188

[36] Ray, D. (1988), Comparison of Forecasts: An Empirical Investigation,*Sankhya: The Indian Journal of Statistics,* Vol 50, Series B, part 2, 258-277

[37] Satchell, S. and A. Timmermann (1993), An Assessment of the Economic Value of Nonlinear Foreign Exchange Rates Forecasts, *Discussion Paper 6192, Birkbeck College.*

[38] Scheinkman, J. A. and B. Le Baron (1989), Non-Linear Dynamics and Stock Returns, *J of Business*, 64, 311-338

[39] Taylor, S. J. (1982), Tests of the Random Walk Hypothesis Against a Price Trend Hypothesis, *J of Financial and Quantitative Analysis*, 17, 37-61

[40] Taylor, S. J. (1980), Conjectured Models for Trends in Financial Prices, Tests and Forecasts, *J of the Royal Statistical Society*, A, 143, 338-362

[41] Tiao, G. C. and R. S. Tsay (1994), Some Advances in Non-Linear and Adaptive Modelling in Time-Series, *J of Forecasting*, Vol 13, 109-131

[42] Tong, H. (1978), On a Threshold Model, in *Pattern Recognition and Signal Processing*, ed. C.H. Chen, Amsterdam: Sijhoff and Noordhoff

[43] Tong, H. (1983), *Threshold Models in Non-Linear Times Series Analysis*, Vol 21 of Lecture Notes in Statistics, ed. K. Krickegerg, NY, Springer-Verlag

[44] Tong, H. (1990) *Nonlinear Time Series*, Oxford Statistical Science Series 6, Oxford University Press, NY

[45] Tong, H. and K. S. Lim (1980), Threshold Autoregression, Limit Cycles, and Cyclical Data (with discussion), *J. R. Statistical Society*, Series B, 42, 245-292

[46] Tsay, R. S. (1989), Testing and Modelling Threshold Autoregressive Process, *J of American Statistical Association*, Vol 84, No 405, 231-240

[47] Westerfield, J. M. (1977), An Examination of Foreign Exchange Risk under Fixed and Floating Regimes, *J of International Economics*, 7, 181-200